Visco-TTI-Elastic FWI using Discontinuous Galerkin

Curtis C. Ober, Thomas M. Smith, James R. Overfelt, S. Scott Collis, Gregory J. von Winckel, Bart G. van Bloemen Waanders, Nathan J. Downey, Scott A. Mitchell, Stephen D. Bond, David F. Aldridge, Sandia National Laboratories*
Jerome R. Krebs ExxonMobil Upstream Research Company

SUMMARY

The need to better represent the material properties within the earth's interior has driven the development of higher-fidelity physics, e.g., visco-tilted-transversely-isotropic (visco-TTI) elastic media and material interfaces, such as the ocean bottom and salt boundaries. This is especially true for full waveform inversion (FWI), where one would like to reproduce the real-world effects and invert on unprocessed raw data. Here we present a numerical formulation using a Discontinuous Galerkin (DG) finite-element (FE) method, which incorporates the desired high-fidelity physics and material interfaces. To offset the additional costs of this material representation, we include a variety of techniques (e.g., non-conformal meshing, and local polynomial refinement), which reduce the overall costs with little effect on the solution accuracy.

INTRODUCTION

There are two primary tasks to obtain and improve the characterization of the subsurface through Full Waveform Inversion (FWI): increase the fidelity of the physics (e.g., governing equations and material representation), and reduce the computational costs to determine the earth's material parameters. Obviously higher-fidelity physics, such as elasticity, attenuation and anisotropy, are needed to accurately simulate wave propagation and mimic real-world conditions and this requires additional computational resources.

Another important aspect of this higher-fidelity physics is the spatial representation of the material properties, including material discontinuities (e.g., sediments, faults, ocean bottom, and salt bodies). Finite-Difference (FD) schemes have difficulty representing these discontinuities, while the class of finite-element schemes (e.g., continuous finite element, spectral elements, and Discontinuous Galerkin) can naturally represent the discontinuities across the element interfaces and use unstructured grids to track them across the domain.

Respecting the above high-fidelity requirements, we have selected Discontinuous Galerkin schemes which are a natural choice for the first-order hyperbolic equations commonly used in FWI today. Continuous FE and spectral schemes are well-suited for elliptic equations (Hesthaven and Warburton, 2008, p. 7), but usually require some sort of stabilization for hyperbolic equations. All these schemes can reduce the computational costs by employing higher-order methods and h/p adaptivity, but DG schemes can easily incorporate non-conformal and hybrid meshes through the flux calculations and Riemann solves. Additionally, continuous FE and spectral schemes are implicit and require the mass matrix to be inverted. For transient problems, this becomes a disadvantage.

For forward modeling, DG methods have been applied to a variety of seismic physics (acoustic/elastic interfaces (Wilcox et al., 2010), elasticity (Käser and Dumbser, 2006), attenuation (Käser et al., 2007), anisotropy (de la Puente et al., 2007) and poro-elasticity (de la Puente, 2008)), with h/p adaptivity (Etienne et al. (2010), Mercerat and Glinsky (2015)), non-conformal (Mazzieri et al., 2013), and hybrid meshes.

Full Waveform Inversion (FWI) imposes additional computational requirements to determine the adjoints and gradients. FWI has been demonstrated on a variety of seismic physics including acoustic and elastic with attenuation (Gholami et al. (2010), Peter et al. (2011), Wilcox et al. (2014)). Here we show an implementation for performing FWI on visco-TTI-elastic media using the DG formulation. We also show results demonstrating some of these capabilities.

FORMULATION AND METHODOLOGY

The FWI problem can be formulated as a constrained optimization problem with a least-squares objective function, $\min_{\Phi} J(\mathbf{U},\Phi)$, where

$$J \equiv \frac{1}{2}\sum_{r=1}^{N_r}\int_{\mathsf{Q}} \xi_r(\mathbf{x})\ \left\|\mathsf{R}\left(\mathbf{U}-\hat{\mathbf{U}}\right)\right\|^2 d\mathsf{Q} + \frac{\beta}{2}\int_{\Omega}\left\|\mathsf{R}\left(\Phi\right)\right\|^2 d\Omega \tag{1}$$

subject to the visco-TTI-elastic equations

$$\frac{\partial \sigma_I}{\partial t} - \tilde{C}_{IJ}\partial V_J - \sum_{\ell=1}^{L} r_I^{\ell} = \frac{\partial m_I^s}{\partial t} \tag{2}$$

$$\frac{\partial v_i}{\partial t} - \nu\frac{\partial \sigma_{ij}}{\partial x_j} = \nu\left[f_i + \frac{\partial m_{ij}^a}{\partial x_j}\right] \tag{3}$$

$$\frac{\partial r_I^{\ell}}{\partial t} + a_{\ell}\omega_{\ell}\tilde{C}_{IJ}\partial V_J + \omega_{\ell} r_I^{\ell} = 0 \tag{4}$$

over the computational domain Ω, and time horizon T, with $\mathsf{Q} = \Omega \times (0,T]$. In these expressions, σ_I and σ_{ij} are the stresses in Voigt and standard tensor notation respectively; $\tilde{C}_{IJ}$ is the rotated stiffness matrix; r_I^{ℓ} are the memory variables for attenuation; $m_{ij} = m_{ij}^s + m_{ij}^a$ are the symmetric and anti-symmetric parts of the moment density tensor; v_i are the particle velocities; ν is the specific volume; f_i is the force density; a_{ℓ} are the relaxation amplitudes for a single quality factor, Q; ω_{ℓ} are the relaxation frequencies; and

$$\partial V = \left\{ \begin{array}{c} \partial v_x/\partial x \\ \partial v_y/\partial y \\ \partial v_z/\partial z \\ \partial v_y/\partial z + \partial v_z/\partial y \\ \partial v_x/\partial z + \partial v_z/\partial x \\ \partial v_x/\partial y + \partial v_y/\partial x \end{array} \right\}. \tag{5}$$

The material properties are parameterized in several ways throughout the code. The user often supplies the material properties in *primitive* parameters, $\Phi' = (\rho, V_p, V_s, \delta, \varepsilon, \gamma, Q, \theta, \phi)^T$. Here ρ is density, V_p is the compressional wave speed; V_s is the shear wave speed; $\delta, \varepsilon, \gamma$ are the Thomsen (1986) parameters; Q is the quality factor; and θ, ϕ are the material rotation angles. The governing equations are expressed in *native* parameters, $\tilde{\Phi} = (\nu, \tilde{C}_{IJ}, a_\ell)^T$, and the unrotated version $\Phi = (\nu, C_{IJ}, a_\ell)^T$, where C_{IJ} is the orthorhombic stiffness matrix. The optimization algorithms operate with the *optimization* parameters, which can be Φ' or Φ, or some other parameterization (e.g., impedance, $I_p = \rho V_p$; slowness, $S_p = V_p^{-1}$; and wave speed ratio, $V_{sp} = V_s/V_p$.). These transformations between parameterizations are nonlinear and require contraction operations (e.g., 21 $\tilde{C}_{IJ}$ coefficients to 5 primitive parameters, and L relaxation amplitudes a_ℓ to one quality factor Q).

The relaxation amplitudes and frequencies are determined from the quality factor, Q, using the least-squares approach of Emmerich and Korn (1987):

$$\sum_{\ell=1}^{L} \frac{Q(\omega_k)\omega_k\omega_\ell + \omega_\ell^2}{\omega_k^2 + \omega_\ell^2} a_\ell = 1. \tag{6}$$

Here we have L relaxation mechanisms, and have created an over-determined system of k discretization frequencies with $1 \le k \le K$ and $K = 2L-1$. The relaxation frequencies are specified in the range $[\omega_{min}, \omega_{max}]$ by

$$\omega_\ell = \begin{cases} \sqrt{\omega_{min}\omega_{max}} & \text{for } L = 1 \\ \omega_{min}^{(1-\xi_\ell)}\, \omega_{max}^{\xi_\ell} & \text{for } L > 1 \end{cases} \tag{7}$$

where $\xi_\ell = (\ell - 1)/(L-1)$.

Discontinuous Galerkin Formulation

We can write the governing equations in the strong form as

$$\mathbf{U}_{,t} + \mathbf{A}_i\mathbf{U}_{,i} + \mathbf{R}\mathbf{U} = \mathbf{S} \tag{8}$$

where $\mathbf{U}$ is the vector of state variables, $\mathbf{A}_i$ are Jacobian matrices, $\mathbf{R}$ is the relaxation matrix, and $\mathbf{S}$ is the sources. To obtain the DG formulation, we construct the weak form over the elements in Ω, apply integration by parts, introduce the numerical flux, and apply reverse integration by parts to get

$$\begin{aligned} \textstyle\sum_{e=1}^{N} \int_{\Omega_e} \mathbf{V} \cdot \left(\mathbf{U}_{,t} + \mathbf{A}_i\mathbf{U}_{,i} + \mathbf{R}\mathbf{U} + \mathbf{S}\right) d\Omega \\ + \int_{\partial\Omega_e} \mathbf{V} \cdot \left(\hat{\mathbf{F}}_n - \mathbf{F}_n\right) d\Gamma = 0 \end{aligned} \tag{9}$$

where $\mathbf{V}$ is the weighting function, $\hat{\mathbf{F}}_n = \hat{\mathbf{F}}_n(\mathbf{U}^-, \mathbf{U}^+)$ is the numerical flux which depends on the left $\mathbf{U}^-$ and right $\mathbf{U}^+$ states, $\mathbf{F}_n = \mathbf{A}_n\mathbf{U}$ is the real flux, and $\mathbf{A}_n$ is the Jacobian matrix oriented in the face normal direction n_i. Reverse integration by parts removes the requirement that $\mathbf{A}_i$ be differentiable.

The numerical flux connects the elements by solving the Riemann problem across the element faces. It also provides numerical stability through *upwinding* and satisfies the consistency condition, $\hat{\mathbf{F}}_n(\mathbf{U}, \mathbf{U}) = \mathbf{F}_n$. Two examples are Lax-Friedrichs (Toro, 1997, p. 172) and Steger and Warming (1981),

$$\begin{aligned} \hat{\mathbf{F}}_n^{LF} &= \tfrac{1}{2}\left[\mathbf{A}_n\mathbf{U}^- + \mathbf{A}_n\mathbf{U}^+ + |\mathbf{A}_n^{LF}|(\mathbf{U}^- - \mathbf{U}^+)\right] \\ \hat{\mathbf{F}}_n^{SW} &= \tfrac{1}{2}\left[\mathbf{A}_n\mathbf{U}^- + \mathbf{A}_n\mathbf{U}^+ + |\mathbf{A}_n^{SW}|(\mathbf{U}^- - \mathbf{U}^+)\right] \end{aligned} \tag{10}$$

where $|\mathbf{A}_n^{LF}| = |\lambda_{max}|\mathbf{I}$ is a diagonal matrix scaled by the maximum eigenvalue (wave speed). The Steger-Warming dissipation matrix $|\mathbf{A}_n^{SW}| = \mathscr{R}|\Lambda|\mathscr{R}^{-1}$ is constructed from the eigen-decomposition of $\mathbf{A}_n$, where $\mathscr{R}$ and $\mathscr{R}^{-1}$ are the right and left eigenvectors respectively, and $|\Lambda|$ is a diagonal matrix containing the absolute value of the individual eigenvalues.

Boundary Conditions

Boundary conditions are derived from the kinematics of the equations. At interfaces between two solid media, (-) and (+), the velocities $\mathbf{v}^- = \mathbf{v}^+$ and traction $\sigma^- \cdot \mathbf{n} = \sigma^+ \cdot \mathbf{n}$ are continuous. These conditions are enforced automatically in the DG formulation. For the interface between a solid (-) and fluid (+), the normal component of velocity is continuous $\mathbf{v}^- \cdot \mathbf{n} = \mathbf{v}^+ \cdot \mathbf{n}$ and the tangential component of the traction is zero. At a free interface between a solid or liquid (-) and a vacuum (+), $\sigma^- \cdot \mathbf{n} = 0$. Finally, it is useful to define a *non-reflecting* boundary condition that allow waves to leave the domain, and is specified by setting the outside state to $\mathbf{U}^+ = 0$.

Spatial Discretization

Referring back to Eq. (9), the computational domain is subdivided into N elements, quadrilaterals or triangles in two-dimensions, hexahedrals or tetrahedrals in three-dimensions. Legendre polynomials form the basis functions for spatial discretization. Two separate DG formulations are considered. The first is referred to as *modal* because solution variables are transformed into modes of polynomials through projection operations. Volume integrals are evaluated using a Gauss-Labatto-Legendre (GLL) quadrature rule. Surface integration uses a Gauss-Legendre (GL) quadrature rule where the GL points are interpolated from the GLL points. In the implementation of the modal formulation, an arbitrary number of quadrature points (q) can be specified for each element and exact integration of any term in Eq. (9) can be achieved. The GLL points also produce a diagonal mass matrix for affine elements. In addition, the media is projected into the space of Legendre polynomials which are then interpolated to the GLL or GL integration points. The polynomial order representation of the media is independent of that used for the state variables.

The second formulation is referred to as *spectral* due to its resemblance to the spectral element method (Karniadakis and Sherwin (1999)). This formulation is very similar to the modal formulation. The main differences are that solution variables remain in physical space, and volume and surface integrals are approximated with the GLL points. Therefore, the number of quadrature points for both state and media are constrained to be $q = p + 1$. This results in under-integration, however the error decreases as p increases. The spectral formulation has several advantages compared to the modal formulation. First, no projection/interpolation of the solution variables or the media is required except at the interface of two elements with different polynomial orders; this includes hanging nodes. This results in a factor of two reduction in computation time. Second, the mass matrix is always diagonal even for non-affine elements. This can reduce the computation time by a factor of 5 to 10 depending on the polynomial order.

Non-conformal Faces and Element Orders

In order to represent strong material variations, slow moving

surface waves (e.g., Scholte waves) and source and receiver locations, the solution mesh can be non-conformal with "hanging nodes." We allow a 1 to N refinement across cell boundaries: a single face S of one element can have N neighboring faces $\{s_i\}$. However, the neighbor faces must be a partition of the larger face; i.e., $s_i \subseteq S$ and s_i cannot overlap both S and some other face S'. We call such s_i and S *non-conformal* faces.

The $\{s_i\}$ are not required to be the same shape but can vary in both size and aspect ratios. There is no restriction on the element size and local level of refinement. Given one hanging node mesh, elements can be arbitrarily refined again, as long as $s_i \subseteq S$ is maintained everywhere.

A *boxtree* mesh is our generalization of the common octree, where each parent can be subdivided into $x \times y \times z$ children. We have a couple of extensions to align the mesh with material discontinuities. First, if the domain cuts boxtree hexes into trivalent polyhedra then midpoint subdivision can divide them into hexes (Li et al., 1997). Second, one may completely remove boxes cut by, or near, the discontinuity; then mesh that space using tetrahedra; while connecting the two meshes with non-conformal faces. We utilize the VoroCrust method for surface reconstruction using Voronoi cells (Mahmoud et al., 2015). It is very robust over complicated domains, and can generate either polyhedra or their dual tetrahedra. The face flux calculations of the DG method provides great mesh flexibility, allowing irregular non-conforming faces with different types (e.g., tetrahedra and hexes) and orders.

In both hybrid and hex meshes, there are no restrictions on the relative element orders between S and $\{s_i\}$, and it is often useful to use lower-order elements on the refined side. The flux across non-conformal faces is conserved by ensuring the total flux, as well as the moments up to the order of the side with fewer orders, are conserved across the interface. These meshing capabilities allow an overall reduction in the number of degrees of freedom, and thus computational cost.

Local Polynomial Refinement

The polynomial order within each element can be selected for the shortest wavelength, $\lambda_{min} = c_{min}/f_{max}$, expected within the element. Here c_{min} is the slowest wave speed (e.g., V_p, V_s or Scholte wave speed) and f_{max} is the maximum expected frequency. Note f_{max} may be substantially smaller in elements deep in the model, as higher frequencies attenuate. Using lower order provides a substantial computational savings. With a sufficient number of points per wavelength, $\alpha \approx 5$, and side length h, we can choose the local polynomial order p:

$$p = \alpha\, h\, f_{max}/c_{min} - 1. \tag{11}$$

Time Integrators

To integrate Eqs. (9, 13, 14), we have a variety of integrators available, e.g., trapezoidal rule, four-stage Runge-Kutta, and a low-storage 14-stage Runge-Kutta (LSRK14) (Niegemann et al., 2012). The LSRK14 has the potential to reduce the costs by 2x because of the increased stability region, and decrease the memory requirements for the adjoint solve by 10x.

Adjoint Equations

To determine the adjoint and material gradients, we form a discrete Lagrange functional

$$\begin{aligned}\mathscr{L}(\mathbf{U},\mathbf{W},\Phi) = J(\mathbf{U},\Phi) + \textstyle\int_{\mathsf{P}_e} \mathbf{W}\cdot\left(\hat{\mathbf{F}}_n - \mathbf{F}_n\right) d\mathsf{P}_e \\ + \textstyle\sum_{e=1}^{N_e} \int_{\mathsf{Q}_e} \mathbf{W}\cdot\left(\mathbf{U}_{,t} + \mathbf{A}_i\mathbf{U}_{,i} + \mathbf{R}\mathbf{U} - \mathbf{S}\right) d\mathsf{Q}_e\end{aligned} \tag{12}$$

where $\mathbf{W}$ is the adjoint variable (i.e., the Lagrange multiplier). We can obtain the state equation, Eq. (8), by holding Eq. (12) stationary with respect to $\mathbf{W}$, $\partial_{\mathbf{W}}\{\mathscr{L}(\mathbf{U},\mathbf{W},\Phi)\} = 0$. To obtain the adjoint equation, we hold Eq. (12) stationary with respect to $\mathbf{U}$,

$$\begin{aligned}\textstyle\sum_{e=1}^{N_e} \int_{\mathsf{Q}_e} \Big(-\mathbf{V}\cdot\mathbf{W}_{,t} - \mathbf{V}\cdot\mathbf{A}_i^T\mathbf{W}_{,i} + \mathbf{V}\cdot\mathbf{R}^T\mathbf{W} \\ + \mathbf{V}\cdot\textstyle\sum_{r=1}^{N_r} \xi_r(\mathbf{x})\, \mathsf{R}^T\mathsf{R}\left(\mathbf{U} - \hat{\mathbf{U}}\right)\Big) d\mathsf{Q}_e \\ + \mathbf{V}\cdot\partial_{\mathbf{U}}\left\{\textstyle\int_{\mathsf{P}_e}\left(\mathbf{W}\cdot\hat{\mathbf{F}}_n\right) d\mathsf{P}_e\right\} = 0\end{aligned} \tag{13}$$

Note this equation is solved for the adjoint variable, $\mathbf{W}$, and that the source term is replaced with the difference between the state and measured data, and integrated backward in time. Also the last term needs to be found for each numerical flux (e.g., $\hat{\mathbf{F}}_n^{LF}$ and $\hat{\mathbf{F}}_n^{SW}$).

To solve for $\mathbf{W}$ requires the state $\mathbf{U}$ at all time steps to be available in order for Eq. (13) to be integrated back in time. But the state data for all time steps can be prohibitively large to save. In this case, the Griewank Algorithm (Restrepo et al. (1998)) can be used, where a select subset of states are stored and intermediate states are recomputed.

Gradient Equations

Lastly the gradient equation is found by holding Eq. (12) stationary with respect to Φ.

$$\begin{aligned}\partial_\Phi\mathscr{L} = \beta\textstyle\int_\Omega \Phi\, d\Omega + \sum_{e=1}^{N_e}\int_{\mathsf{Q}_e}\left(\mathbf{W}\cdot\frac{\partial\mathbf{A}_i}{\partial\Phi}\mathbf{U}_{,i} - \mathbf{W}\cdot\frac{\partial\mathbf{S}}{\partial\Phi}\right) d\mathsf{Q}_e \\ + \partial_\Phi\left\{\textstyle\sum_{e=1}^{N_e}\int_{\mathsf{P}_e}\mathbf{W}\cdot\left(\hat{\mathbf{F}}_n - \mathbf{F}_n\right) d\mathsf{P}_e\right\} = 0\end{aligned} \tag{14}$$

The regularization term is present, and again the last term needs to be determined for each numerical flux.

INVERSION

We have several optimization packages available, including an internal linesearch capability, and the Rapid Optimization Library (ROL). ROL (Kouri et al., 2014) is Trilinos package (Heroux and Willenbring, 2012) currently under development, designed for efficient large-scale numerical optimization. ROL provides the capability to solve both full-space and reduced-space problems. For FWI, the storage requirements for a three-dimensional inversion makes full-space methods prohibitively expensive and therefore we use the reduced-space framework. Currently, ROL is used to solve FWI problems using a variety of Linesearch and Trust-Region methods.

In the Linesearch setting, descent directions can be taken to be the negative gradient (steepest descent), generated by nonlinear conjugate gradient or secant methods, or an inexact Newton method. In the latter case, the approximate action of the Hessian on a direction vector v

$$\frac{\nabla f(x+\varepsilon v) - \nabla f(x)}{\varepsilon} = [\nabla^2 f(x)]v + O(\varepsilon) \tag{15}$$

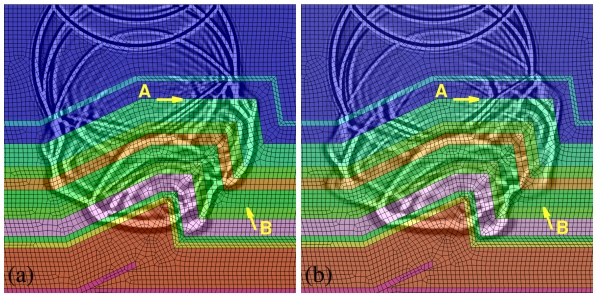

Figure 1: Wave propagation through an elastic overthrust model without a) and with b) attenuation and anisotropic material properties.

and the conjugate-gradient method is applied to approximately solve the Newton step vector. ROL selects the Linesearch step length based on a number of methods for generating candidate step lengths, such as backtracking, Brent's method, cubic interpolation, in conjunction with a variety of sufficient decrease and curvature requirements which can be set by the user. Similarly, the finite-difference Hessian is used to generate local quadratic models in Trust-Region methods.

One can include simultaneous source inversion (SSI), Krebs et al. (2009), in Eqs. (1)-(3) by replacing the receiver data with the receiver data of multiple shots, N_s, and the source with the sources of multiple shots,

$$\hat{\mathbf{U}} \rightarrow \sum_{s=1}^{N_s} \omega_s \hat{\mathbf{U}}_s \qquad m_{ij} \rightarrow \sum_{s=1}^{N_s} \omega_s (m_{ij})_s \tag{16}$$

where ω_s is randomly 1 or -1 based on s, and is changed every optimization iteration. Note that SSI is independent of the optimization package, and for optimization algorithms with history, e.g., conjugate-gradient methods, the history is corrupted by the randomization of the SSI method.

SYNTHETIC RESULTS

In Fig. 1, we have wave propagation through an overthrust models without a) and with b) attenuation and anisotropic material properties. The material properties vary over the entire domain (not shown). Throughout the images, one can find subtle differences in the wave propagation due to the higher-fidelity physics provided by visco-TTI-elastic media. Point A indicates a difference in wave propagation due to anisotropic material properties (i.e., bent waves and additional reflected waves). Point B indicates differences due to attenuation (i.e., a reduction in wave amplitude).

To demonstrate our inversion capabilities, we display a sample of the results from Krebs et al. (2016) in Fig. 2. In this inversion of I_p, V_p and V_s from visco-TTI-elastic data, we are just showing the results related to V_p. For further details, see Krebs et al. (2016). In Fig. 2(a), the target V_p values are shown with an anomaly in the form of the letter "P". There are other anomalies in other material properties, e.g., "I" for impedance, "S" for V_s, and "q" for the quality factor (all not shown).

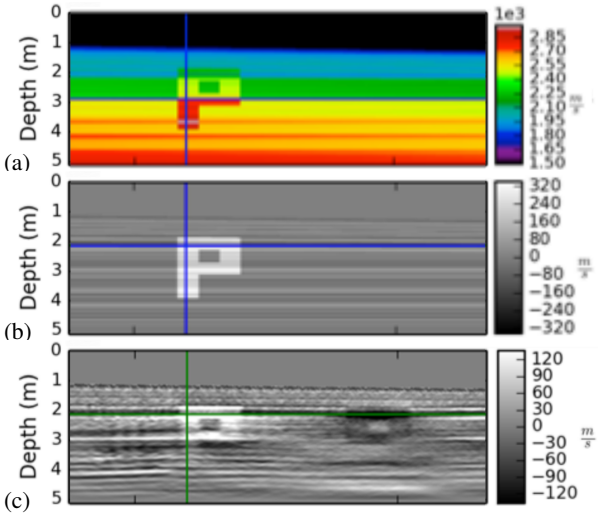

Figure 2: FWI results from Krebs et al. (2016) for (a) the target model for V_p, (b) the expected update to V_p, and (c) the inverted results for V_p.

Fig. 2(b) shows the update (target model minus initial model), which we hope to achieve in our inversions for V_p, and Fig. 2(c) shows our inverted results after 430 iterations. The inverted V_p is good, but there is cross-talk from the impedance (just to the left of the "P"), and the quality factor (the "q" to the far right) which are due to the lack of updates to the quality factor. Note that there is little to no cross-talk from V_s (between the "P" and "q"). Although there are open questions related to the cross-talk between parameters, the inverted results demonstrate the working capabilities of our formulation.

DISCUSSION

We have presented a numerical formulation for full waveform inversion (FWI) of a visco-TTI-elastic media using a Discontinuous Galerkin (DG) method. The DG method provides a flexible means to accurately represent material interfaces, while allowing a variety meshing capabilities that reduces the number of degrees of freedom and thus the computational costs. Additionally, we have shown examples of wave propagation and inversion with further details shown in (Krebs et al., 2016) that demonstrate the ability to model and invert high-fidelity physics. This capability should allow FWI on raw-data, and reduce the need to preprocess field data.

ACKNOWLEDGMENTS

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

REFERENCES

de la Puente, J., 2008, Seismic wave simulation for complex rheologies on unstructured meshes: PhD thesis, Ludwig-Maximilians-Universität.

de la Puente, J., M. Käser, M. Dumbser, and H. Igel, 2007, An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes-IV. anisotropy attenuation: Geophysical Journal International, **169**, 1210–1228.

Emmerich, H., and M. Korn, 1987, Incorporation of attenuation into time-domain computations of seismic wave fields: Geophysics, **52**, 1252–1264.

Etienne, V., E. Chaljub, J. Virieux, and N. Glinsky, 2010, An hp-adaptive discontinuous galerkin finite-element method for 3-d elastic wave modelling: Geophysical Journal International, **183**, 941–962.

Gholami, Y., A. Ribodetti, R. Brossier, S. Operto, and J. Virieux, 2010, Sensitivity analysis of full waveform inversion in vti media: Presented at the 72nd EAGE Conference and Exhibition incorporating SPE EUROPEC 2010, SPE EAGE.

Heroux, M., and J. Willenbring, 2012, A new overview of the trilinos project: Scientific Programming, **20**, 83–88.

Hesthaven, J. S., and T. Warburton, 2008, Nodal discontinuous Galerkin methods: Alorithms, analysis and applications: Springer, volume **54** *of* Texts in Applied Mathematics.

Karniadakis, G. E., and S. J. Sherwin, 1999, Spectral/*hp* element methods for CFD: Oxford University Press.

Käser, M., and M. Dumbser, 2006, An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes -I. The two-dimensional isotropic case with external source terms: Geophysical Journal International, **166**, 855–877.

Käser, M., M. Dumbser, J. de la Puente, and H. Igel, 2007, An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes-III. Viscoelastic attenuation: Geophysical Journal International, **168**, 224–242.

Kouri, D., D. Ridzal, B. von Bloemen Waanders, and G. von Winckel, 2014, Rapid optimization library: Technical Report SAND2014-19572, Sandia National Laboratories, Albuquerque, NM.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full wave seismic inversion using encoded sources: Geophysics, **74**, 177–188.

Krebs, J. R., C. C. Ober, T. M. Smith, J. R. Overfelt, S. S. Collis, G. J. von Winckel, B. G. van Bloemen Waanders, N. J. Downey, and D. Aldridge, 2016, Synthetic study of raw-data fwi applied to visco-tti-elastic data: SEG Technical Program Expanded Abstracts. (Submitted).

Li, T., C. Armstrong, and R. McKeag, 1997, Quad mesh generation for k-sided faces and hex mesh generation for trivalent polyhedra: Finite Elements in Analysis and Design, **26**, 279 – 301.

Mahmoud, A. H., A. A. Rushdi, S. A. Mitchell, A. A. Abdelrazek, C. L. Bajaj, J. D. Owens, and M. S. Ebeida, 2015, VoroCrust: Simultaneous surface reconstruction and volume meshing with Voronoi cells: http://www.poems15.gatech.edu/schedule. (Talk slides for the workshop Polytopal Element Methods in Mathematics and Engineering, Georgia Tech. Organizers Chunmei Wang, Andrew Gillette, and Lourenco Beirão da Veiga.).

Mazzieri, I., M. Stupazzini, R. Guidotti, and C. Smerzini, 2013, Speed: Spectral elements in elastodynamics with discontinuous galerkin: a non-conforming approach for 3d multi-scale problems: International Journal for Numerical Methods in Engineering, **95**, 991–1010.

Mercerat, E. D., and N. Glinsky, 2015, A nodal high-order discontinuous Galerkin method for elastic wave propagation in arbitrary heterogeneous media: Geophysical Journal International, **201**, 1101–1118.

Niegemann, J., R. Diehl, and K. Busch, 2012, Efficient low-storage runge-kutta schemes with optimized stability regions: J. Comput. Phys., **231**, 364–372.

Peter, D., D. Komatitsch, Y. Luo, R. Martin, N. Le Goff, E. Casarotti, P. Le Loher, F. Magnoni, Q. Liu, C. Blitz, T. Nissen-Meyer, P. Basini, and J. Tromp, 2011, Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes: Geophysical Journal International, **186**, 721–739.

Restrepo, J. M., G. K. Leaf, and A. Griewank, 1998, Circumventing storage limitations in variational data assimilation studies: SIAM Journal of Scientific Computing, **19**, 1586–1605.

Steger, J. L., and R. F. Warming, 1981, Flux Vector Splitting of the Inviscid Gasdynamics Equations with Application to Finite-Difference Methods: Journal of Computational Physics, **40**, 263–293.

Thomsen, L., 1986, Weak elastic anisotropy: Geophysics, **51**, 1954–1966.

Toro, E. F., 1997, Riemann solvers and numerical methods for fluid dynamics : A practical introduction: Springer.

Wilcox, L., G. Stadler, C. Burstedde, and O. Ghattas, 2010, A high-order discontinuous Galerkin method for wave propagation through coupled elastic-acoustic media: Journal of Computational Physics, **229**, 9373–9396.

Wilcox, L. C., G. Stadler, T. Bui-Thanh, and O. Ghattas, 2014, Discretely exact derivatives for hyperbolic pde-constrained optimization problems discretized by the discontinuous galerkin method: Journal of Scientific Computing, **63**, 138–162.